NREM3 Sleep Stage Estimation Based on Accelerometer by Body Movement Count and Biological Rhythms

Daiki Shintani[1*], **Iko Nakari**[1*], **Satomi Washizaki**[2], **Keiki Takadama**[1*]

[1]The University of Electro-Communications, 1-5-1 Chofugaoka, Chofu, Tokyo, Japan, 182-8585
[2]Brain Sleep Co., LTD JP Tower 26F, Marunouchi 2-7-2, Chiyoda-ku, Tokyo, Japan, 100-7026
shindaiki@cas.lab.uec.ac.jp, iko0528@cas.lab.uec.ac.jp, satomi.washizaki@brain-sleep.com, keiki@inf.uec.ac.jp

Abstract

This paper proposes the method by physiological knowledge to improve the estimation performance of the NREM3 sleep based on the waist-attached accelerometer. Specifically, this paper proposes the hybrid method that combines the method based on body movement counts and the method based on biological rhythms of sleep. Through the human subject experiment, the following implications were revealed: (1) the proposed method can outperform famous machine learning models (Random Forest and LSTM) trained with automatically generated features that do not sufficiently incorporate domain knowledge; (2) when the input features are based on domain knowledge, the estimator explicitly designed by humans can outperform the machine learning method; and (3) combining the body movement counting method and the biological rhythm-based method can suppress the error of the body movement counting method and reduce false positives.

Introduction

Sleep has an impact on the individual and social well-being (Lowe, Safati, and Hall 2017). Extensive research reports the harmful effects of sleep deprivation. For example, short sleep duration was reported to correlate with adverse effects on individual physical health (Chami et al. 2020), such as obesity, type II diabetes, cardiovascular disease, and mortality, as well as individual mental health, such as emotional balance, life satisfaction, and depression (Pilcher, Ginter, and Sadowsky 1997). Such effects on individual as well as social well-being have also been reported, affecting increased government healthcare costs, population safety, and social performance (Perez-Pozuelo et al. 2020). Despite these facts, the average sleep duration worldwide tends to be decreasing (Bonnet and Arand 1995). In particular, Japan has the shortest sleep duration of all OECD member countries, one and a half hours less than the global average, and it is estimated that many of its citizens are suffering from sleep deprivation (Organization for Economic Cooperation and Development 2021). Given such trends and their impact on well-being, it is beneficial to monitor sleep conditions in daily life to detect sleep problems at an early stage.

The clinical method for measuring sleep quality (sleep stages) is polysomnography (PSG) based on the Rechtschaffen & Kales (R&K) method (Rechtschaffen and Kales 1968), but it requires manual scoring by the specialist and is physically and mentally stressful because multiple electrodes are attached to the head and face. In recent years, with the development of AI and IT technologies, sleep stage estimation methods employing sensors that are less burdensome to measure in daily life, such as mattress sensors (Nakari and Takadama 2023), radio signals(Zhao et al. 2017), and accelerometers, have been attracting attention. In particular, sleep measurement by accelerometers has been employed in many studies for a long time because it does not require a special environment (Boe et al. 2019; Sundararajan et al. 2021; Gu et al. 2014). However, most studies employing accelerometers have focused on sleep onset estimation, and machine learning methods that simply learn typical statistics calculated from accelerometers have poor performance (Sundararajan et al. 2021). This is because data obtained from accelerometers have the following problems compared to that from PSG tests: (1) because the information from the accelerometer is only body movements, it is difficult to separate sleep stages from typical statistics; and (2) the amount of data is not sufficient at the experimental stage, making it difficult to learn an effective estimator.

To improve the accuracy of sleep stage estimation based on accelerometers, this paper proposes the estimation method based on physiological knowledge for the NREM3 sleep. Specifically, this paper proposes the estimation method for the NREM3 sleep that manually combines decisions based on body movement counts and biological rhythms during NREM3 sleep. Furthermore, this paper shows experimentally that the proposed method, which incorporates physiological domain knowledge of sleep, achieves higher performance than machine learning methods that learn automatically generated statistics.

Sleep Mechanism

Sleep Stage

Sleep stage is an indicator of the depth of human sleep as defined by the R&K method (Rechtschaffen and Kales 1968). The sleep stages are, from shallowest to highest: the WAKE, REM, Non-REM1 (NREM1), NREM2, NREM3,

[*]These authors contributed equally.

and NREM4. Following the American Academy of Sleep Medicine (AASM) scoring manual (Berry et al. 2012), the NREM3 and NREM4 sleep are combined into one stage (the NREM3 sleep) in this paper. The R&K method acquires three biological data during sleep, consisting of electrooculogram (EEG), electromyogram (EMG), and electrooculogram (EOG) during sleep. The expert determines the sleep stage at 30-second intervals from the data.

Ultradian Rhythm

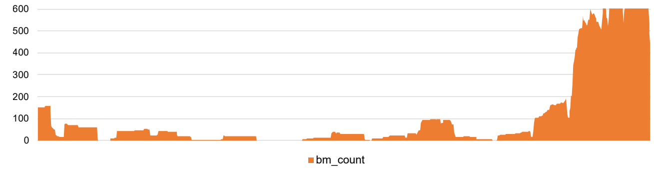
Sleep is a temporal process with a sequence of sleep stages and dependencies between sequential epochs (Phan et al. 2018). The ultradian rhythm is the sleep cycle of 60 to 120 minutes consisting of shallow sleep (NREM1, NREM2, REM) and deep sleep (NREM3). A standard sleep time of approximately 7 hours and 30 minutes is associated with 4-5 sleep cycles. It is known that the sleep cycle tends to be deeper just after bedtime than before waking. In general, this rhythm changes according to the rhythm of daily life and physical characteristics so that even the same person does not maintain the same sleep state and rhythm. This fact indicates that the identification of ultradian rhythms should be conducted every night.

NREM3 Features

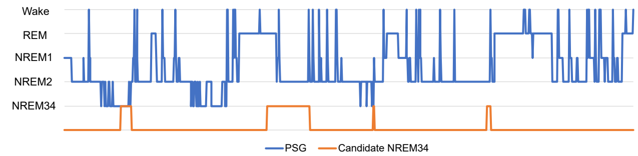
The NREM3 sleep is known as the deep sleep stage and accounts for 15-25% of total sleep time. During this stage, the brain produces delta waves, which are slow brain waves of 2 Hz or less, making it difficult to awaken. The heart rate and respiratory rate decrease, and the brain and body are in resting mode in the NREM3 sleep. The characteristics of the NREM3 sleep employed in accelerometer-based estimation are as follows: (i) large body movements are rarely observed; (ii) the NREM3 sleep often appears according to the ultradian rhythm cycle; and (iii) the periods are rarely too short (less than 8 minutes) or too long (more than 40 minutes). This paper designs the estimation method based on the domain knowledge of these sleep patterns.

Realted Works

Sleep stage estimation by AI has been studied to address the problem of requiring an expert to score sleep stages from data collected by the PSG test. Classical sleep stage estimation was based on human knowledge of sleep physiology to build features and estimation models (Virkkala et al. 2007). In recent years, the methods that learn from automatically generated features (Van Der Donckt et al. 2023) and deep learning methods that do not require feature design (Eldele et al. 2021; Perslev et al. 2019) have achieved high accuracy without explicitly incorporating domain knowledge. This is due to the availability of relatively large PSG datasets containing hundreds of whole-night records and advances in machine learning and deep learning techniques. However, machine learning models trained with general time series features for accelerometers have been reported to perform poorly except for WAKE stage estimation. This paper investigates the effectiveness of a manual method that explicitly incorporates domain knowledge.



(a) Example of the body movement count.



(b) Example of the estimation from the body movement count.

Figure 1: The overview of NREM3 estimation based on body movement counts.

Proposeed Method

Bodymovement Count Based NREM3 Estimation

Figure 1 shows an overview of the NREM3 estimation based on body movement counts. The NREM3 sleep is estimated based on the physiological feature that body movements are not observed for a certain period of time, which flow is summarized as follows.

1. To focus only on the magnitude of the body motion, the absolute translation value $V_{abs,i} = |V_{n,i} - 1|$ is calculated for the norm vector $V_{n,i}$ of the accelerometer of the epoch i.
2. The value is set to 0 in order to ignore the effect of small body motion. The mean value of all $V_{abs,i}$ in step 1. from $i = 0$ to n is denoted by $mean$ and the standard deviation by σ, as shown in equation(1).

$$V'_{abs,i} = \begin{cases} 0 & \text{if } V_{abs,i} < mean - 0.8 \times \sigma \\ V_{abs,i} & \text{otherwise} \end{cases} \tag{1}$$

3. The number of large body movements exceeding a certain threshold in the past 25-minute window is counted for the value $V'_{abs,i}$ of step 2. e $V'_{abs,i}$ of $V'_{abs,i}$. Figure 1a shows an example of body movement count. Let $mean'$ and σ' be the mean and standard deviation of $V'_{abs,i}$, respectively, then the body movement count $N_{count,i}$ is calculated by Eq. (2).

$$N_{count,i+1} = \begin{cases} N_{count,i} + 1 & \text{if } V'_{abs,i-50:i} > \\ & mean' + 4 \times \sigma' \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

4. The epoch with a body movement counts $N_{count,i}$ of 0 in step 3. is determined to be the NREM3 sleep. Figure 1b shows an example of NREM3 estimated from the body movement count feature.

Ultradian-Rhythm Based NREM3 Estimation

Figure 2 shows an overview of NREM3 estimation based on ultradian rhythms (UR). The UR-based method identifies the UR, which is the cycle of approximately 90-minute

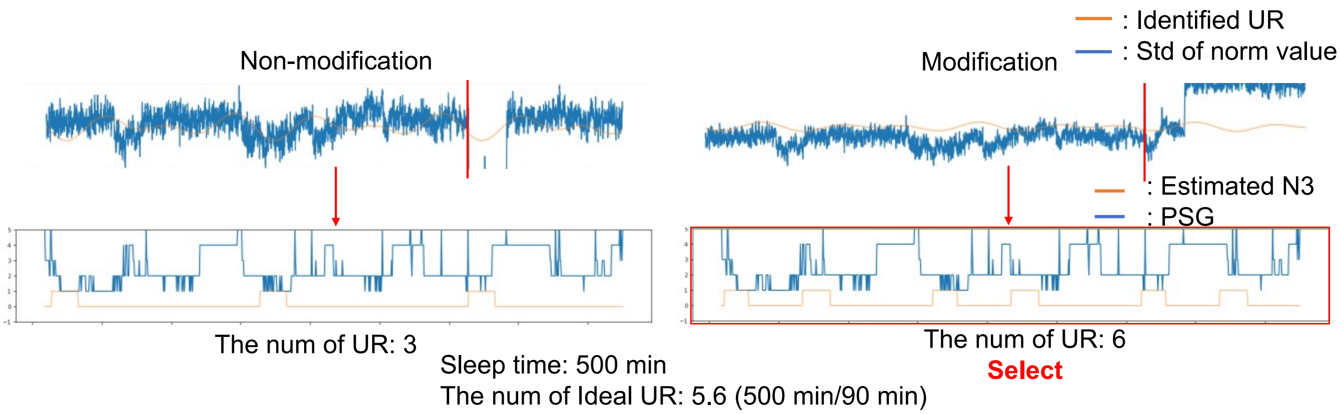


Figure 2: The overview of NREM3 estimation based on ultradian rhythm.

sleep stages, and determines the period when the estimated UR curve falls as the NREM3 sleep. To reduce the effect of large changes in posture on the identification of UR, the correction process is selected by comparing the case with and without the correction process that aligns the average values before and after the posture change. The estimation method is summarized as follows.

1. After calculating the standard deviation in a 1-second window for the normed values of acceleration, a moving average is calculated in a 30-second window for further smoothing.
2. The regression model $f(t)$ of the trigonometric function is fitted by the least-squares method to the norm values in step 1. The regression model $f(t)$ synthesizes frequency waves as shown in Eq. (3), where a_l and b_l are the coefficients of the cos/sin wave with period L, L is set to $\{90, 60\}$ minutes, and C is the constant term in $f(t)$. Note that a_l, $b_l(l \in L)$, and C are optimized by the least-squares method, where a_l and $b_l(l \in L)$ are two times the standard deviation of norm values and C is initialized to 1.

$$\begin{aligned} f(t) &= \Sigma_{l \in L}\{a_l \cos m_l t + b_l \sin m_l t\} + C \\ m_l &= \frac{2\pi}{l} \end{aligned} \tag{3}$$

3. When the z-axis value changes significantly, the norm value also changes before and after the change, so the regression model f(t) is fitted after correcting for this effect. Specifically, for the values in step 1, the later values are modified so that the average value of the 20 minutes before and after a large change in the z-axis value is equal to the former value. The regression model $f(t)'$ is the same as in Eq. (3) as in step 2.
4. The function without (step 2) and with (step 3) modification by z-axis change is selected if the number of periods of NREM3 sleep is closer to the number of UR periods assumed from the total sleep time. The total sleep time T divided by the average UR period of 90 minutes is compared to the number of NREM3 sleep periods estimated without and with z-axis modification. The function $f_{auto}(t)$ is chosen in Eq. (4), where N and N' are the number of NREM3 sleep periods estimated without and with z-axis modification.

$$f_{auto}(t) = \begin{cases} f(t) & \text{if } |T/90 - N| \leq |T/90 - N'| \\ f(t)' & \text{otherwise} \end{cases} \tag{4}$$

5. The 30-secondly estimates output from the $f_{auto}(t)$ function are normalized by [0, 1] and are estimated to be NREM3 when the value is less than 0.25.

Hybrid Method

The hybrid method integrates the body movement count-based and the UR-based estimation methods for the NREM3 sleep. The method extends or eliminates NREM3 based on the relationship between UR-based estimation and body movement count estimation, which is summarized as follows.

1. If the estimation of the body movement count base overlaps with the estimation of the UR base by more than 1/3, the NREM3 sleep is expanded to include both the body movement count base and the UR base. However, if the expanded NREM3 is continuous for more than 40 minutes, the common part of the body movement count base and the UR base are used as the NREM3 sleep.
2. If the body movement count-based estimation overlaps with the UR-based estimation by less than 1/3, it is employed for the NREM3 sleep.
3. If the UR-based estimate of the body movement count-based estimate does not overlap but there is a UR-based estimate within 8 minutes of the body movement count-based estimate, then the NREM3 sleep is used for the body movement count-based estimate.
4. If the body-movement count-based and UR-based estimates do not overlap and there is no UR-based estimate within 8 minutes of the body-movement count-based estimate, the NREM3 sleep is removed.

Dataset and Evaluation

Dataset

The dataset consists of 29 whole-night sensor data from the 20s to 50s. In this paper, a coin-shaped sensor (BRAIN SLEEP COIN, Brain Sleep Co. Ltd.) attached to a nightwear around waist is employed as an acceleration sensor. The coin-shaped sensor measured the norm data of acceleration with a sampling rate of 10 Hz, z-axis data of acceleration with a sampling rate of 1 Hz (1: face up, 0: sideways, -1: face down), and temperature with a sampling rate of 1 Hz. After sleep, the sleep stage, according to the R&K method, is determined every 30 seconds from EEG, EOG, and EMG and is employed as the target label. For this experiment, the ethics community of Ota General Hospital approved this study in agreement with Helsinki's declaration.

Evalutation

All experiments employ the leave-one-out cross-validation, where the model is trained on 28 nights of data and evaluated on the other night. The evaluation indicators are accuracy, precision, recall, f1-measure, and specificity between the correct sleep stage of the PSG test and the estimated sleep stage.

Ex. 1: Is It Necessary To Incorporate Domain Knowledge Into AI?

Experimental Setting

Experiment 1 compares the proposed system that incorporates sleep domain knowledge with machine learning methods that do not adequately account for sleep domain knowledge. The proposed method is compared with the following methods:

- RF with automatically generated features

  Random Forest (RF) (Breiman 2001) is an ensemble learning algorithm composed of decision trees as weak classifiers. RF determines its output by majority voting on multiple trees. The parameters of RF are set as follows: (i) the maximum depth of decision trees is 10; and (ii) the number of decision trees is 100. The features to be input to RF are computed using the automatic feature generation and selection method (Nakari and Takadama 2023), which generates a large number of time series features and selects valid features for the U-test. This method is employed to compute the features that are valid for estimating the NREM3 stages. To deal with class imbalance, WAKE, REM, NREM12, and NREM3 are adjusted to have the same proportion of each.

- LSTM with automatically generated features

  LSTM (Graves and Graves 2012) is the recurrent neural network model that stores long-term dependencies. The parameters of LSTM are as follows: (i) the batch size is 128; (ii) Adam (Kingma and Ba 2014) is employed as the optimization method with the learning rate of 0.01; and (iii) the number of epochs is 50. The input features are the same as for RF.

Experimental Result

Figure 3 shows the boxplots for accuracy, precision, recall, f1-measure, and specificity. Blue, orange, and gray represent the results of RF, LSTM, and the proposed method, respectively. The figure shows that the proposed method achieves higher average values than the other ML methods for all indicators. In particular, the result shows accuracy and specificity are less variation than the other two methods.

Figure 4a, 4b, and 4c show the results of sleep stages estimated by RF, LSTM, and proposed method, respectively, for a certain night data. The blue lines are the correct sleep stages of the PSG test in order of shallow sleep: WAKE, REM, NREM1, NREM2, and NREM3. The orange lines are the results of the NREM3 sleep estimated by each method; when the line is on the NREM3, it is estimated to be the NREM3 sleep, and when the line is below the NREM3, it is not the NREM3 sleep. Figure 4a shows that the estimation results by RF change in a few epochs, which is contrary to the finding that sleep stages do not change frequently. Figure 4b shows that LSTM is able to estimate NREM3 continuously, but there are many areas where NREM3 is estimated where NREM3 is not. Figure 4c shows that NREM3 is estimated continuously, and there are a few areas where NREM3 is incorrectly estimated where it is not NREM3.

Ex. 2: Design Estimator by ML vs Human

Experimental Setting

Experiment 2 compares the performance of the proposed hybrid method with that of the ML model (RF) estimation of NREM3 from two sets of features, body-movement count-based and UR-based. The ML-based method learns the RF using the counts in step 3 of the Bodymovement count based NREM3 estimation and the outputs of the regression model normalized in steps 2 and 3 of the UR based NREM3 estimation section as features. The parameters of RF are set as follows: (i) the maximum depth of decision trees is set to 3; and (ii) the number of decision trees is set to 100. The proposed hybrid method is the same as Experiment 1.

Experimental Result

Figure 5 shows the results of estimating NREM3 from body motion count-based and UR-based features using the ML method and human-designed rules. The blue bars show the results of estimating NREM3 with the ML method (RF), and the orange bars show the results of estimating NREM3 with the proposed hybrid method. The results show that the human-designed hybrid method performs better than the ML method on all evaluation indicators.

Disccusion

Effectiveness of Estimation Based on Domain Knowledge

Figure 3 shows that the proposed method designed based on domain knowledge of sleep achieves better performance than machine learning models (RF and LSTM) trained on

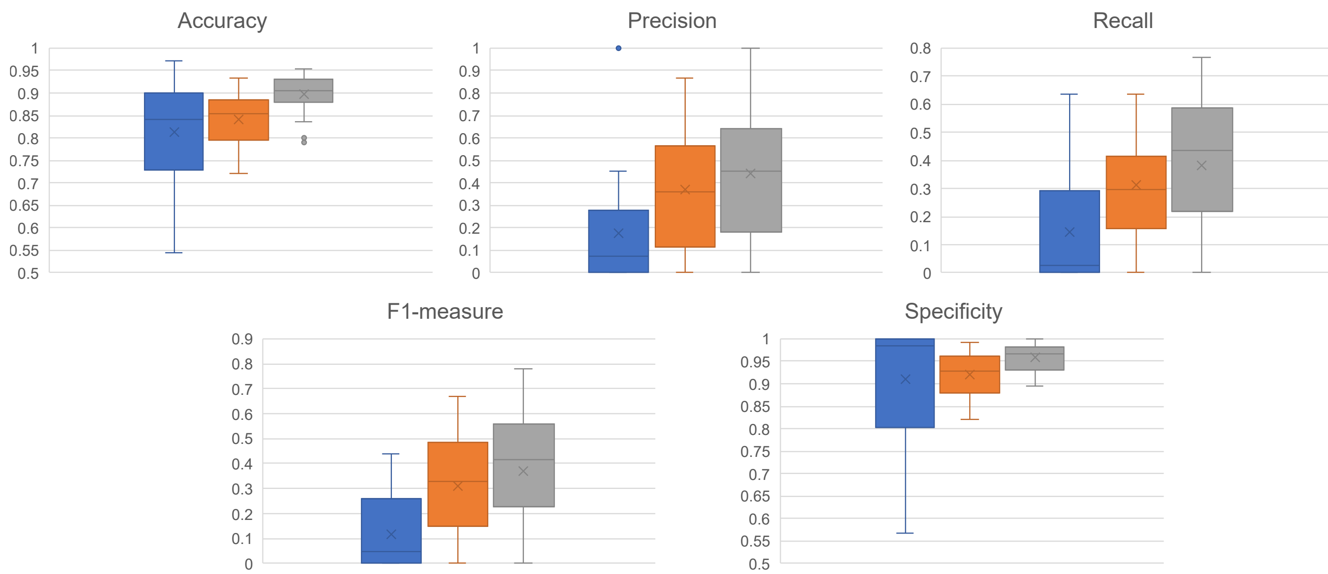


Figure 3: The results of NREM3 estimation.

automatically generated time series features. The small variation in accuracy and specificity implies that the explicit incorporation of physiological knowledge suppressed obvious errors. The same trend can be seen in the results of Figure 4a, 4b, and 4c. Even when designing the estimator, the results in Figure 5 suggest that the proposed hybrid method, which explicitly incorporates knowledge of the physiology of sleep, is effective. Therefore, these results suggest the importance of incorporating domain knowledge into the design, rather than simply applying machine learning to the task.

Ablation Study

To investigate the effectiveness of the proposed method, the performance of the body movement count-based, the UR-based, and the hybrid estimation methods are compared. Figure 6 shows the result of each indicator for the body movement count-based, UR-based, and hybrid methods. Blue bars indicate the results of the body movement count-based estimation, orange bars indicate the results of the UR-based estimation, and gray bars indicate the results of the hybrid method estimation. The figure shows that the body movement count-based and hybrid methods perform better than the UR-based method on all evaluation indicators. The hybrid method performs worse than the body movement count-based method on Recall and F1-measure, but better on Precision and Specificity. This is because the hybrid method takes into account physiological rhythms in the UR-based method in addition to the body-movement count-based method, and thus eliminates the misclassification of the body movement count-based method that does not fit the rhythms. This suggests that the hybrid method can reduce false positives by compensating for false positives in the body movement count base.

Conclusion

This paper proposed the hybrid method combining the body movement count-based estimation and the UR-based estimation designed from the physiological characteristics of sleep to estimate sleep stages from accelerometers attached to the nightwear around waist in daily life, focusing on the NREM3 sleep. Through experiments on human subjects, the proposed method showed higher performance in all evaluation indicators compared to machine learning methods with automatic feature generation. The hybrid method incorporating physiological sleep knowledge into the count-based and UR-based features of body movements outperformed the ML method. These results indicate that incorporating domain knowledge into feature design and estimation decisions is effective in improving performance when the quantity and quality of the dataset are relatively poor. A limitation of this paper is that the experiments were conducted on the limited dataset and tasks. It needs to be confirmed whether similar trends can be obtained with a wider range of data sets and tasks.

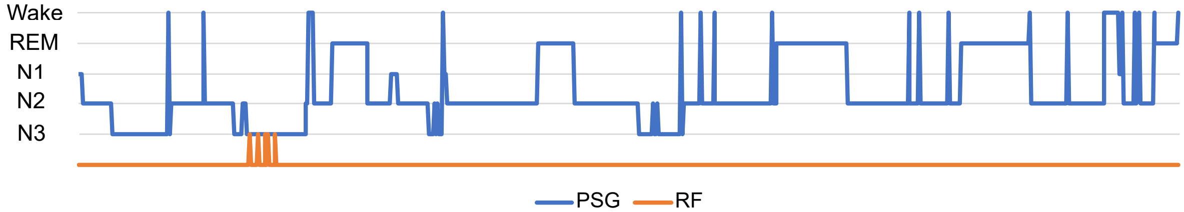
Acknowledgments

This work was supported by Brain Sleep Co. LTD.

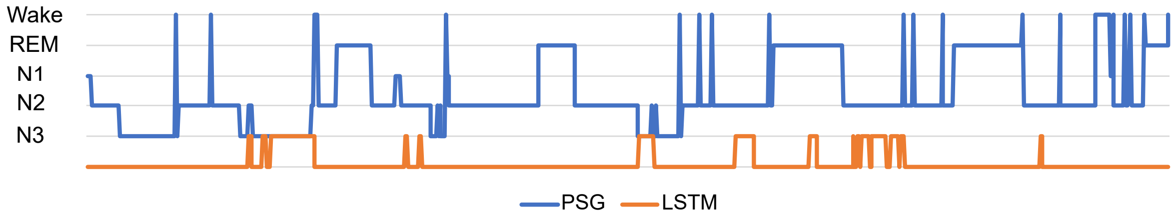
References

Berry, R. B.; Brooks, R.; Gamaldo, C. E.; Harding, S. M.; Marcus, C.; Vaughn, B. V.; et al. 2012. The AASM manual for the scoring of sleep and associated events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*, 176: 2012.

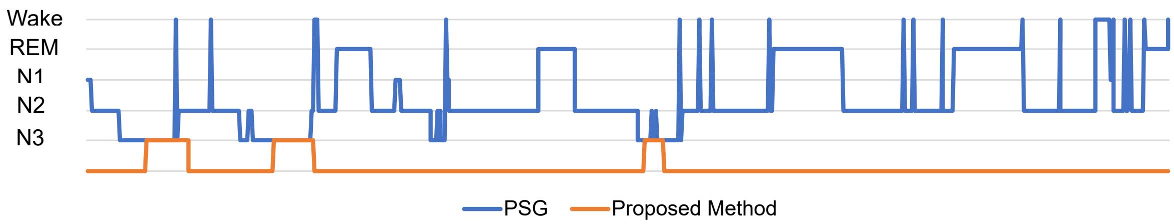
Boe, A. J.; McGee Koch, L. L.; O'Brien, M. K.; Shawen, N.; Rogers, J. A.; Lieber, R. L.; Reid, K. J.; Zee, P. C.; and Jayaraman, A. 2019. Automating sleep stage classification us-



(a) The detail of the result of RF.



(b) The detail of the result of LSTM.



(c) The detail of the result of Proposed Method.

Figure 4: The detailed sleep stage.

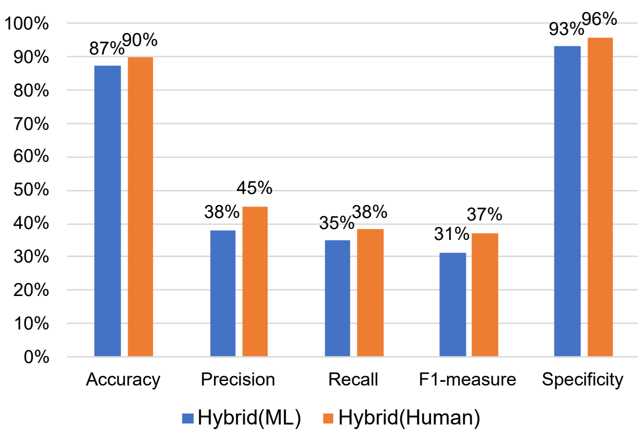


Figure 5: The results of Ex. 2.

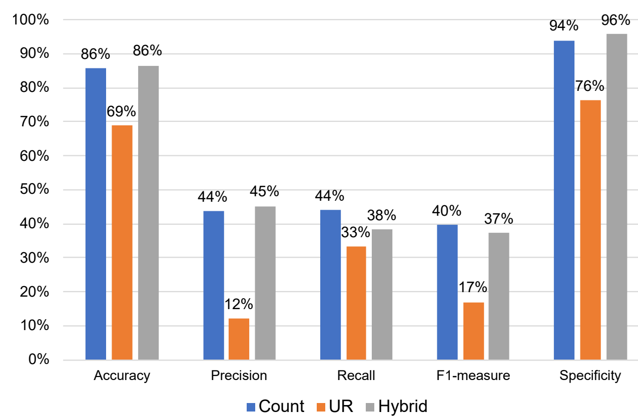


Figure 6: The results of Ex. 3.

ing wireless, wearable sensors. *NPJ digital medicine*, 2(1): 131.

Bonnet, M. H.; and Arand, D. L. 1995. We are chronically sleep deprived. *Sleep*, 18(10): 908–911.

Breiman, L. 2001. Random forests. *Machine learning*, 45: 5–32.

Chami, H. A.; Ghandour, B.; Isma'eel, H.; Nasreddine, L.; Nasrallah, M.; and Tamim, H. 2020. Sleepless in Beirut: sleep duration and associated subjective sleep insufficiency, daytime fatigue, and sleep debt in an urban environment. *Sleep and Breathing*, 24: 357–367.

Eldele, E.; Chen, Z.; Liu, C.; Wu, M.; Kwoh, C.-K.; Li, X.; and Guan, C. 2021. An attention-based deep learning approach for sleep stage classification with single-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29: 809–818.

Graves, A.; and Graves, A. 2012. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, 37–45.

Gu, W.; Yang, Z.; Shangguan, L.; Sun, W.; Jin, K.; and Liu, Y. 2014. Intelligent sleep stage mining service with smartphones. In *Proceedings of the 2014 ACM international Joint Conference on pervasive and ubiquitous Computing*, 649–660.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lowe, C. J.; Safati, A.; and Hall, P. A. 2017. The neurocognitive consequences of sleep restriction: a meta-analytic review. *Neuroscience & Biobehavioral Reviews*, 80: 586–604.

Nakari, I.; and Takadama, K. 2023. Personalized Non-contact Sleep Stage Estimation with Weighted Probability Estimation by Ultradian Rhythm. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1–4. IEEE.

Organization for Economic Cooperation and Development. 2021. GENDER EQUALITY, Gender Data Portal. https://www.oecd.org/gender/data/. Accessed: 2024-1-9.

Perez-Pozuelo, I.; Zhai, B.; Palotti, J.; Mall, R.; Aupetit, M.; Garcia-Gomez, J. M.; Taheri, S.; Guan, Y.; and Fernandez-Luque, L. 2020. The future of sleep health: a data-driven revolution in sleep science and medicine. *NPJ digital medicine*, 3(1): 42.

Perslev, M.; Jensen, M.; Darkner, S.; Jennum, P. J.; and Igel, C. 2019. U-time: A fully convolutional network for time series segmentation applied to sleep staging. *Advances in Neural Information Processing Systems*, 32.

Phan, H.; Andreotti, F.; Cooray, N.; Chén, O. Y.; and De Vos, M. 2018. Joint classification and prediction CNN framework for automatic sleep stage classification. *IEEE Transactions on Biomedical Engineering*, 66(5): 1285–1296.

Pilcher, J. J.; Ginter, D. R.; and Sadowsky, B. 1997. Sleep quality versus sleep quantity: relationships between sleep and measures of health, well-being and sleepiness in college students. *Journal of psychosomatic research*, 42(6): 583–596.

Rechtschaffen, A.; and Kales, A. 1968. *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects*. Washington DC.

Sundararajan, K.; Georgievska, S.; Te Lindert, B. H.; Gehrman, P. R.; Ramautar, J.; Mazzotti, D. R.; Sabia, S.; Weedon, M. N.; van Someren, E. J.; Ridder, L.; et al. 2021. Sleep classification from wrist-worn accelerometer data using random forests. *Scientific reports*, 11(1): 24.

Van Der Donckt, J.; Van Der Donckt, J.; Deprost, E.; Vandenbussche, N.; Rademaker, M.; Vandewiele, G.; and Van Hoecke, S. 2023. Do not sleep on traditional machine learning: Simple and interpretable techniques are competitive to deep learning for sleep scoring. *Biomedical Signal Processing and Control*, 81: 104429.

Virkkala, J.; Hasan, J.; Värri, A.; Himanen, S.-L.; and Müller, K. 2007. Automatic sleep stage classification using two-channel electro-oculography. *Journal of neuroscience methods*, 166(1): 109–115.

Zhao, M.; Yue, S.; Katabi, D.; Jaakkola, T. S.; and Bianchi, M. T. 2017. Learning sleep stages from radio signals: A conditional adversarial architecture. In *International Conference on Machine Learning*, 4100–4109. PMLR.